Performance of Quadratic Time-Frequency Distributions as Instantaneous Frequency Estimators

Veselin N. Ivanović, Miloš Daković, LJubiša Stanković

Abstract— **General performance analysis of the shift covariant class of quadratic time-frequency distributions (TFDs) as instantaneous frequency (IF) estimators, for an arbitrary frequency-modulated (FM) signal, is presented. Expressions for the estimation bias and variance are derived. This class of distributions behaves as an unbiased estimator in the case of monocomponent signals with a linear IF. However, when the IF is not a linear function of time, then the estimate is biased. Cases of white stationary and white nonstationary additive noises are considered. The well-known results for the Wigner distribution (WD) and linear FM signal, and the spectrogram of signals whose IF may be considered as a constant within the lag window, are presented as special cases. In addition, we have derived the variance expression for the spectrogram of a linear FM signal that is quite simple but highly signal dependent. This signal is considered in the cases of other commonly used distributions, such as the Born-Jordan and the Choi-Williams distributions. It has been shown that the reduced interference distributions outperform the WD but only in the case when the IF is constant or its variations are small. Analysis is extended to the IF estimation of signal components in the case of multicomponent signals. All theoretical results are statistically confirmed.**

I. Introduction

Instantaneous frequency (IF) estimation is an important research topic in signal analysis [2], [3], [14]-[22], [28]-[29]. There are several approaches to this problem. Time-frequency distribution (TFD)-based approach is one of them [14]-[16], [18], [28], [29]. The basis for using TFDs in the IF estimation is their first moment property, [3], [4], [12]. The first-order TFD moment, with respect to frequency, provides an acceptable IF definition for a time-varying signal. The TFD, which is used to recover the IF as its first moment, provides an unbiased estimate. The presence of noise, however, leads to a serious degradation of the first moment estimate due to the absence of any averaging in its definition. In other words, the first moment may have a high statistical variance, even for high values of input signal-to-noise ratio (SNR) [22]. TFDs concentrate the energy of a signal at and around the IF in the time-frequency plane, [3], [18], [22], [23]. Consequently, the peak detection of the TFDs is used as an IF estimator, as a natural alternative to the first moment.

The IF estimation based on TFDs maxima is analyzed in [3], [5], [14]-[17], [21], [22], [28], and [29]. Out of the quadratic class of TFDs, only the most frequently used ones are considered there: the Wigner distribution (WD) for linear frequency-modulated (FM) signal and the spectrogram for signals with constant frequency. It has been shown that in the case of noisy signals, this estimate highly depends on the SNR, as well as on the window width.

In this paper, we present a general analysis of an arbitrary shift covariant quadratic TFD as an IF estimator for any FM signal. Expressions for the bias and variance of this IF estimator are derived. When the IF is a nonlinear function of time, then its estimate is biased for all TFDs from this class (Cohen class of TFDs-CD), whereas they behave as unbiased estimators in the case of monocomponent signals with linear IF. The exact expressions for the IF estimator variance in the cases of white stationary and white nonstationary noises are derived. The corresponding expressions for some frequently used TFDs from the CD are obtained as special cases as well. We have presented the well-known re-

IEEE Transactions on Signal Processing, Vol.51, No.1, Jan. 2003.

sults for the pseudo WD and linear FM signal and for the spectrogram of signals whose IF may be considered as a constant. In addition, we have derived the variance expression for the spectrogram of a linear FM signal. This signal is considered in the cases of other commonly used TFDs, such as Born-Jordan (BJD) and Choi-Williams distribution (CWD). It has been shown that the reduced interference distributions (RID) outperform the pseudo WD but only in the case when the IF is constant or its variations are small. For highly nonstationary signals, the pseudo WD can produce better results. The analysis is extended to the multicomponent signals. It has been shown that the results obtained for monocomponent signals remain valid for the multicomponent ones when TFDs from RID class are used and signal components are well separated. For the pseudo WD-based IF estimation, the variance of each component depends on the total power of all signal components.

The paper is organized as follows. After this introduction, the IF estimator is defined, and the problem is described. In Section III, analysis of the estimation error is performed. The bias and variance of the estimation error in the cases of commonly used quadratic TFDs are derived in Section IV. The IF estimation of multicomponent signals is considered in Section V. The obtained results are checked numerically and statistically in Section VI.

II. Background Theory

Consider discrete-time observations

$$x(nT) = f(nT) + \varepsilon(nT),$$
$$f(t) = A(t)\exp(j\phi(t)) \tag{1}$$

where n is an integer, T is a sampling interval, $\varepsilon(nT)$ is a white noise, and $A(t)$ is a slow-varying amplitude of the analyzed signal. By definition, [6], [16], [18], the IF is a first derivative of the signal phase

$$\omega(t) = \phi'(t) \equiv \frac{d\phi(t)}{dt}. \tag{2}$$

Assume that $\omega(t)$ is an arbitrary smooth differentiable function of time with bounded derivatives $\left|\omega^{(r)}(t)\right| = \left|\phi^{(r+1)}(t)\right| \leq S_r(t)$, $r \geq 1$.

The general form of the quadratic shift-covariant TFDs (CD) is, [4], [6], [12], [13], [30]

$$C_x(t,\omega;c) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(\theta,\tau)$$
$$\times x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{-j\theta(t-u)-j\omega\tau}d\theta du d\tau. \tag{3}$$

Its discrete-time domain form is

$$C_x(t,\omega;c) = \frac{T^2}{\pi}\sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty} c(\theta,2nT)$$
$$\times x(kT+nT)x^*(kT-nT)e^{-j\theta(t-kT)-j2\omega nT}d\theta. \tag{4}$$

After the integration over θ, and substitution $kT - t = mT$, we get

$$C_x(t,\omega;\varphi) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\varphi(mT,nT)$$
$$\times x(t+mT+nT)x^*(t+mT-nT)e^{-j2\omega nT} \tag{5}$$

with

$$\varphi(mT,nT) = \frac{T^2}{\pi}\int_{-\infty}^{\infty} c(\theta,2nT)e^{j\theta mT}d\theta.$$

Suppose that $\varphi(t,\tau)$ has a finite width along time and lag directions $\varphi(t,\tau) = 0$ for $|t| > 1/2$, $|\tau| > 1/2$. This is then the pseudo (windowed) form of a TFD. In numerical realization, finite limits must be used. The notation $\varphi_h(mT,nT) = (T/h)^2\varphi(mT/h,nT/h)$ will be used for a finite width h, $h > 0$, of the kernel in both directions. The finite width h is used in definition of the TFDs from CD in order to localize the estimate. The factor $(T/h)^2$ is used so that the integral (sum) of $\varphi_h(t,\tau)$ over time and lag is h independent. Of course, constant factor $(T/h)^2$ does not influence the IF estimation analysis in any way. Note that for most of commonly used TFDs, the kernel is a symmetric function in both time and lag axes. The kernel, which has been represented in its time-lag ($\varphi(t,\tau)$) or ambiguity domain ($c(\theta,\tau)$), determines the TFD characteristics. Different kernels produce different TFDs [4], [12], such as the WD, together with its pseudo and smoothed forms, spectrogram, and the TFDs from the RID class [4], [12], [13], [23].

Let us analyze an arbitrary quadratic TFD of the signal $f(t)$. Using the fact that the signal has a slow-varying amplitude $f(t+mT\pm nT)\varphi_h(mT,nT) \cong A(t)\exp[j\phi(t+mT\pm nT)]\varphi_h(mT,nT)$, and expanding $\phi(t+mT\pm nT)$ into a Taylor series around t (up to the third order term in order to enable one to perform the estimation error analysis in the case of FM signals whose IF is not constant), we get

$$C_f(t,\omega;\varphi_h) = |A(t)|^2 \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} \varphi_h(mT,nT)\times e^{-j[2(\omega-\phi'(t))(nT)-2\phi^{(2)}(t)(mT)(nT)-\triangle\phi(t,mT,nT)]} \quad (6)$$

where $\triangle\phi(t,mT,nT)$ is a residue of the phase. It can be represented as

$$\triangle\phi(t,mT,nT) = \sum_{s=3}^{\infty}\frac{\phi^{(s)}(t)}{s!}[(mT+nT)^s-(mT-nT)^s] = \sum_{s=3}^{\infty}\frac{\phi^{(s)}(t)}{s!}\sum_{k=0}^{s}\binom{s}{k}(mT)^{s-k}(nT)^k[1-(-1)^k]. \quad (7)$$

Note that TFDs from the CD would have a maximum at $\omega=\phi'(t)$ if $\phi^{(s)}(t)=0$ for $s\geq 2$. Thus, the IF estimate can be defined as a solution of the following problem, [16], [22]:

$$\widehat{\omega}_h(t) = \arg[\max_{\omega\in Q_\omega}\{C_x(t,\omega;\varphi_h)\}] \quad (8)$$

where $Q_\omega=\{\omega: 0\leq|\omega|<\pi/2T\}$ is a basic frequency interval. The estimation error, produced at a time-instant t, is

$$\triangle\widehat{\omega}_h(t) = \omega(t)-\widehat{\omega}_h(t). \quad (9)$$

III. Analysis of the Estimation Error

Since the estimate of IF $\widehat{\omega}_h(t)$ is defined by the stationary point of $C_x(t,\omega;\varphi_h)$, it is determined by the zero value of $\partial C_x(t,\omega;\varphi_h)/\partial\omega$. Linearization of $\partial C_x(t,\omega;\varphi_h)/\partial\omega=0$ with respect to [16]

1) small estimation error $\triangle\widehat{\omega}_h(t)$;
2) small residual of the phase deviation $\triangle\phi$;
3) noise ε;
4) squared noise ε^2, which gives

$$\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0+\frac{\partial^2 C_x(t,\omega;\varphi_h)}{\partial\omega^2}|_0\triangle\widehat{\omega}_h(t) +\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\triangle\phi}+\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon +\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}=0 \quad (10)$$

where $|_0$ means that the above derivatives are calculated at the point $\omega=\phi'(t)$, $\varepsilon=0$, and $\triangle\phi(t,mT,nT)=0$. The last three terms in (10) determine the variations of $\partial C_x(t,\omega;\varphi_h)/\partial\omega$ caused by $\triangle\phi$, ε, and ε^2, respectively.

The terms from (10) are [16]

$$\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0=0$$

$$\frac{\partial^2 C_x(t,\omega;\varphi_h)}{\partial\omega^2}|_0=-4\,|A(t)|^2\,R_h(t)$$

$$\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\triangle\phi}\cong 2\,|A(t)|^2\,P_h(t)$$

$$\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2} = 2\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\varphi_h(mT,nT)\times\varepsilon(t+mT+nT)\varepsilon^*(t+mT-nT)\times(-jnT)e^{-j2\phi'(t)nT} \quad (11)$$

whereas $\partial C_x(t,\omega;\varphi_h)/\partial\omega|_0\delta_\varepsilon$ will be given separately. The functions $R_h(t)$ and $P_h(t)$ are defined by

$$R_h(t)=\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\varphi_h(mT,nT)(nT)^2\times e^{j2\phi^{(2)}(t)(mT)(nT)} \quad (12)$$

$$P_h(t)=\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\varphi_h(mT,nT)\times\triangle\phi(t,mT,nT)(nT)e^{j2\phi^{(2)}(t)(mT)(nT)}. \quad (13)$$

Note that $\partial C_x(t,\omega;\varphi_h)/\partial\omega|_0=0$ follows from the symmetry of the kernel $\varphi_h(mT,nT)$. Using the notation

$$Q_h=\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon+\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2} \quad (14)$$

we have

$$\triangle\widehat{\omega}_h(t) = \frac{1}{2R_h(t)}(P_h(t) + \frac{Q_h}{2\left|A(t)\right|^2}). \quad (15)$$

In order to get the exact value of the IF estimator variance, the term $\partial C_x(t,\omega;\varphi_h)/\partial\omega|_0\delta_\varepsilon$ will be expressed by using the inner-product form of the CD [1], [8], [24]

$$C_x(t,\omega;\varphi_h) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\widetilde{\varphi}_h(mT,nT)$$

$$\times[x(t+mT)e^{-j\omega mT}][x(t+nT)e^{-j\omega nT}]^* \quad (16)$$

where $\widetilde{\varphi}_h(mT,nT) = \varphi_h((m+n)T/2,(m-n)T/2)$. Consequently

$$\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon$$

$$= j\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\widetilde{\varphi}_h(mT,nT)(n-m)T$$

$$\times[f(t+mT)\varepsilon^*(t+nT) + f^*(t+nT)\varepsilon(t+mT)]$$

$$\times e^{-j\omega(m-n)T}|_0. \quad (17)$$

We can conclude that for the white noise $\varepsilon(nT)$

$$E\left\{\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon\right\} = 0,$$

and

$$E\left\{\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\right\} = 0.$$

Thus, $E\{Q_h\} = 0$. Therefore, the estimation error bias and variance are

$$E\{\triangle\widehat{\omega}_h(t)\} = \frac{P_h(t)}{2R_h(t)} \quad (18)$$

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\} = \frac{\mathrm{var}\{Q_h\}}{16\left|A(t)\right|^4\left|R_h(t)\right|^2} \quad (19)$$

where $R_h(t)$ is defined by (12). By expanding exponential function $\exp(j2\phi^{(2)}(t)(mT)(nT))$ into a power series, we may represent $R_h(t)$ as

$$R_h(t) = \sum_{i=0}^{\infty}\frac{(-1)^i(2\phi^{(2)}(t))^{2i}}{(2i)!}B_h(2i,2i+2) \quad (20)$$

where

$$B_h(k,l) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\varphi_h(mT,nT)(mT)^k(nT)^l \quad (21)$$

are the moments of $\varphi_h(mT,nT)$. From the kernel $\varphi_h(mT,nT)$ symmetry, it follows that the moments $B_h(k,l)$ are different from zero only for even indexes k, l. By using the most significant terms in (20), $R_h(t)$ may be approximated as

$$R_h(t) \cong B_h(0,2) - 2(\phi^{(2)}(t))^2B_h(2,4). \quad (22)$$

Note that when $T \to 0$, we have:

$$B_h(k,l) \to h^{k+l}b_{k,l}$$

$$= h^{k+l}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\varphi(t,\tau)t^k\tau^l dtd\tau. \quad (23)$$

Now, expressions for the bias and variance, given in general case by (18) and (19), will be analyzed.

A. IF Estimation Variance

In the sequel, we will consider the nonstationary, complex-valued, white, Gaussian noise $\varepsilon(nT)$ with the auto-correlation $R_{\varepsilon\varepsilon}(t+mT,t+nT) = I(t+mT)\delta(m-n)$, $I(t) \geq 0$. The stationary noise can be obtained as a special form of the nonstationary one, with $I(t) = \sigma_\varepsilon^2$.

Proposition 1: *Let $\widehat{\omega}_h(t)$ be a solution of (8). For small estimation error and an FM signal $f(t) = A(t)\exp(j\phi(t))$, the IF estimation variance is*

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$$

$$= \frac{2C_I(t,0;\left|\varphi_{h_1}\right|^2) + C_f(t,\phi'(t);\widetilde{\mathbf{\Phi}}_h)}{8\left|A(t)\right|^4\left|R_h(t)\right|^2} \quad (24)$$

where $C_f(t,\phi'(t);\widetilde{\mathbf{\Phi}}_h)$ is a quadratic TFD (with the new kernel $\widetilde{\mathbf{\Phi}}_h = -\widetilde{\mathbf{\Psi}}_h\mathbf{I}_t\widetilde{\mathbf{\Psi}}_h^*$) of the analyzed signal $f(t)$ at its IF, and $\widetilde{\mathbf{\Psi}}_h = \mathbf{A}_{n-m}.*\widetilde{\boldsymbol{\varphi}}_h$. Here, $\widetilde{\boldsymbol{\varphi}}_h$ is a matrix with elements $\widetilde{\varphi}_h(mT,nT)$, whereas $\mathbf{A}_{n-m}$ is a matrix with elements $A(m,n) = n-m$, for $m,n = 1,2,...,N$ (N represents assumed finite limits

for m, n). The operator .$*$ denotes element-by-element matrix multiplication. The matrix $\mathbf{I}_t$ is a diagonal matrix, with $I(t+nT)$ being its elements. In addition, $C_I(t,0;\left|\varphi_{h_1}\right|^2)$ represents a quadratic TFD of $I(t)$ with a new kernel $\left|\varphi_{h_1}(mT,nT)\right|^2$, $\varphi_{h_1}(mT,nT) = \varphi_h(mT,nT)(nT)$, [24].

Special Case: A linear FM signal1 $f(t) = A(t)e^{jat^2/2}$ corrupted by a stationary, complex, white, Gaussian noise, produces the IF independent variance

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\} = \frac{\sigma_\varepsilon^2[2\sigma_\varepsilon^2 W_h + C_f(0,0;-\widetilde{\mathbf{\Psi}}_h^2)]}{8\left|A(t)\right|^4\left|R_h(t)\right|^2} \tag{25}$$

where

$$W_h = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\left|\varphi_h(mT,nT)\right|^2(nT)^2. \tag{26}$$

The proof of Proposition 1 is given in the Appendix.

B. IF Estimation Bias

Proposition 2: *Let $\widehat{\omega}_h(t)$ be a solution of (8). For small estimation error, the IF estimation bias is*

$$E\{\triangle\widehat{\omega}_h(t)\}$$
$$= \frac{1}{R_h(t)}\sum_{i=0}^{\infty}\frac{(-1)^i[2\phi^{(2)}(t)]^{2i}}{(2i)!}\sum_{s=1}^{\infty}\frac{\phi^{(2s+1)}(t)}{(2s+1)!}$$
$$\times\sum_{k=0}^{s}\binom{2s+1}{2k+1}B_h(2s-2k+2i,2k+2i+2) \tag{27}$$

where $R_h(t)$ is given by (20).

Proof of this Proposition follows from definition (13) of $P_h(t)$, and definition (7) of $\triangle\phi(t,mT,nT)$, by using the symmetry of the time-lag kernel $\varphi(t,\tau)$. Note that the bias is defined by (18).

Special Case: Assuming that the IF of analyzed signal is quadratic $\phi(t) = bt^3/3$, we have

$$E\{\triangle\widehat{\omega}_h(t)\} = \frac{1}{2R_h(t)}\sum_{i=0}^{\infty}\frac{(-1)^i[2\phi^{(2)}(t)]^{2i}}{(2i)!}$$
$$\times(B_h(2i+2,2i+2)+\frac{1}{3}B_h(2i,2i+4))\phi^{(3)}(t). \tag{28}$$

The most significant bias terms are obtained for $i = 0, 1$

$$E\{\triangle\widehat{\omega}_h(t)\} \cong \frac{1}{2R_h(t)}[B_h(2,2)+\frac{1}{3}B_h(0,4)$$
$$-2(\phi^{(2)}(t))^2(B_h(4,4)+\frac{1}{3}B_h(2,6))]\phi^{(3)}(t). \tag{29}$$

Since the higher order moments (21) are small, $B_h(4,4)+B_h(2,6)/3$ is relatively small as compared with $B_h(2,2)+B_h(0,4)/3$ and $B_h(2,4)$ is relatively small as compared with $B_h(0,2)$, when $R_h(t) \cong B_h(0,2)$ according to (22), we can write

$$E\{\triangle\widehat{\omega}_h(t)\} \cong \frac{B_h(2,2)+B_h(0,4)/3}{2B_h(0,2)}\phi^{(3)}(t) \tag{30}$$

or

$$|E\{\triangle\widehat{\omega}_h(t)\}| \le S_2\left|\frac{B_h(2,2)+B_h(0,4)/3}{2B_h(0,2)}\right|$$

$$\text{with } S_2 = \sup_t\{\phi^{(3)}(t)\}. \tag{31}$$

IV. Special Cases of Quadratic TFDs

The expressions for variance and bias in the case of any TFD from CD may be obtained as special cases of (24), (25), and (27). Let us write these expressions for the cases of most important and frequently used TFDs.

1) Pseudo WD: For this TFD,

$$\widetilde{\varphi}_h(mT,nT) = w_h(mT)\delta(m+n)w_h(nT),$$

$$R_h(t) = B_h(0,2) = \sum_{n=-\infty}^{\infty}w_{h_1}^2(nT) \to ThM_2^{w^2}$$

where $w_h(nT) = (T/h)w(nT/h)$ is a real-valued and even window function, $w_{h_1}(nT) = w_h(nT)(nT)$, and $M_r^{w^2} = \int_{-1/2}^{1/2}w^2(\tau)\tau^r d\tau$ is the rth moment of the squared window $w^2(\tau)$. Namely, here we use the notation

$$M_r^w = \int_{-1/2}^{1/2}w(\tau)\tau^r d\tau \tag{32}$$

in order to represent the rth moment of the window $w(\tau)$. The rth moments of the window's $w(\tau)$ kth power $w^k(\tau)$ are denoted by $M_r^{w^k}$. Thus, we get

$$\text{var}\{\triangle\widehat{\omega}_h(t)\} = \frac{WD_I(t,0;w_{h_2}) + 2WD_{I,|f|^2}(t,0;w_{h_2})}{4\left|A(t)\right|^4\left|R_h(t)\right|^2} \tag{33}$$

$$\text{bias}(t,h) = \frac{B_h(0,4)}{6B_h(0,2)}\omega^{(2)}(t) = \frac{1}{6}\frac{M_4^{w^2}}{M_2^{w^2}}\omega^{(2)}(t)h^2 \tag{34}$$

where $w_{h_2}(nT) = w_h^2(nT)(nT)$, whereas $WD_{x,y}$ denotes the cross-WD. For the case of a stationary, white, complex, Gaussian noise

$$\text{var}\{\triangle\widehat{\omega}_h(t)\} = \frac{\sigma_\varepsilon^2}{2\left|A(t)\right|^2}(1+\frac{\sigma_\varepsilon^2}{2\left|A(t)\right|^2})W_w\frac{T}{h^3} \tag{35}$$

where

$$W_w = M_2^{w^4}/(M_2^{w^2})^2 \tag{36}$$

is a constant, dependent on the window $w(\tau)$ form, and $M_2^{w^4}$ is the second-order moment of window $w^4(\tau)$. Values of W_w for some commonly used windows are presented in Table I. Note that for the rectangular window $w_h(nT)$, variance (35) and bias (34) are reduced to the well-known expressions from [16]

$$\text{var}\{\triangle\widehat{\omega}_h(t)\} = \frac{6\sigma_\varepsilon^2}{\left|A(t)\right|^2}(1+\frac{\sigma_\varepsilon^2}{2\left|A(t)\right|^2})\frac{T}{h^3} \tag{37}$$

$$\text{bias}(t,h) = \frac{1}{40}\omega^{(2)}(t)h^2. \tag{38}$$

One may conclude that in the case of analyzed FM signals, $\text{var}\{\triangle\widehat{\omega}_h(t)\}$ is not dependent on the phase $\phi(t)$ and its derivatives, i.e., $\text{var}\{\triangle\widehat{\omega}_h(t)\}$ is constant for all values of $\phi^{(2)}(t)$ in the case of linear FM signals.

2) Spectrogram: Here, we have:

$$\widetilde{\varphi}_h(mT,nT) = w_h(mT)w_h(nT).$$

In this case, two parts of variance $\text{var}\{Q_h\}$ [see (68) from the Appendix] have the forms

$$\text{var}\{\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon\} = 2STFT_I(t,0;w_{h_1}^2)SPEC_f(t,\phi'(t);w_h) + 2STFT_I(t,0;w_h^2)SPEC_f(t,\phi'(t);w_{h_1}) - 4STFT_I(t,0;w_{h_2}) \cdot \text{Re}[STFT_f(t,\phi'(t);w_h)STFT_f^*(t,\phi'(t);w_{h_1})] \tag{39}$$

$$\text{var}\{\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\} = 2\,\text{Re}[STFT_I(t,0;w_{h_1}^2)STFT_I^*(t,0;w_h^2)] - 2SPEC_I(t,0;w_{h_2}) \tag{40}$$

where $STFT(t,\omega;w_h)$ represents the short-time Fourier transform, $SPEC(t,\omega;w_h) = \left|STFT(t,\omega;w_h)\right|^2$, whereas $R_h(t)$, (20), and (21), is

$$R_h(t) = \frac{h^2}{4}\sum_{i=0}^{\infty}\frac{(-1)^i(h^2\phi^{(2)}(t)/2)^{2i}}{(2i)!} \times \sum_{i_1=0}^{2i}\sum_{i_2=0}^{2i+2}\binom{2i}{i_1}(-1)^{i_2}\binom{2i+2}{i_2} \times M_{4i+2-i_1-i_2}^w \cdot M_{i_1+i_2}^w. \tag{41}$$

Substitution of (39) and (40) into (68) produces variance $\text{var}\{Q_h\}$. Then, substitution of the obtained variance and (41) into (19) gives the IF estimator variance in the case of the spectrogram. From (39), it can be concluded that $\text{var}\{\triangle\widehat{\omega}_h(t)\}$, in the case of spectrogram, is highly signal dependent. In the same way, starting from (28) and (30), the bias of the IF estimator can be obtained [25].

Special Case: Consider linear FM signal $f(t) = A(t)\exp(jat^2/2)$ corrupted by the stationary [or quasistationary $I(t+nT) = I(t)$], complex, white, Gaussian noise. Then, we have $SPEC_f(0,0;w_{h_1}) = 0$ and $STFT_I(t,0;w_{h_2}) = 0$. Thus

$$\text{var}\{\frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon\} = 2\sigma_\varepsilon^2 ThM_2^{w^2}\cdot SPEC_f(0,0;w_h) \tag{42}$$

where

$$SPEC_f(0,0;w_h) = \left|A(t)\right|^2\left|\sum_{i=0}^{\infty}\frac{(jh^2a/2)^i}{i!}M_{2i}^w\right|^2 \tag{43}$$

TABLE I
COEFICIENTS W_w, S_w, AND C_w, DEFINED BY (36), (45), AND (46), RESPECTIVELY, FOR VARIOUS WINDOW $w(\tau)$ FORMS. TRUNCATED GAUSSIAN WINDOW $w(\tau) = \exp(-(\pi\tau)^2)$ IS CONSIDERED.

Window $w(\tau)$	**Rectangular**	**Hanning**	**Hamming**	**Triangular**	**Gaussian**
W_w	12	54.46	41.66	34.29	36.40
S_w	12	28.11	19.73	19.20	17.89
C_w	$5.3\cdot\ 10^{-3}$	$1.77\cdot\ 10^{-3}$	$2.94\cdot\ 10^{-3}$	$2.74\cdot\ 10^{-3}$	$3.20\cdot\ 10^{-3}$

whereas $R_h(t)$ is given by (41) with $\phi^{(2)}(t) = a$. Now, the exact IF estimation error $\triangle\widehat{\omega}_h(t)$ variance may be obtained by substituting (41)-(43) into (19).

Since the rth moment of the window $w(\tau)$ is a rapidly decreasing function of the order r, then for a relatively small a ($a^2 C_w h^4 < 5$; in the realization from Fig. 1, this means $a \leq 0.6$), $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ can be closely approximated by the following simple form:

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\} \cong \frac{\sigma_\varepsilon^2}{2\left|A(t)\right|^2}\frac{T}{h^3}S_w e^{a^2 C_w h^4} \quad (44)$$

which is obtained by substituting $i = 0, 1$ into (41) and (43). Here

$$S_w = M_2^{w^2}/(M_2^w)^2 \quad (45)$$

and

$$C_w = \frac{1}{4}\left((\frac{M_2^w}{M_0^w})^2 + \frac{M_6^w}{M_2^w} - \frac{M_4^w}{M_0^w}\right) \quad (46)$$

are the window $w(\tau)$-dependent constants; see Table I [25]. Note that due to the kernel $\widetilde{\varphi}_h(mT, nT)$ symmetry, the same values of variance $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ hold for negative a with $a \to |a|$. Conclude that in this case, $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ is not constant. It is highly signal dependent. As a increases, $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ increases from the value

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\} \cong \frac{\sigma_\varepsilon^2}{2\left|A(t)\right|^2}S_w\frac{T}{h^3}\ ,\ \text{for } a = 0 \quad (47)$$

which has been derived in literature as the spectrogram variance [15]. Of course, it holds only for $a = 0$, whereas for other values of a, more general relations (41)-(44), which are derived in this paper, hold.

3) Smoothed Pseudo WD: In this case [4], [6],

$$\varphi_h(mT, nT) = \gamma\exp(-(mT)^2/\alpha - (nT)^2/\beta).$$

For $\alpha = \beta$, $\widetilde{\varphi}_h(mT, nT) = w_h(mT)w_h(nT)$, where $w_h(mT) = \sqrt{\gamma}\exp(-(mT)^2/(2\alpha))$ is the truncated Gaussian window. Consequently, the bias and variance expressions may be directly obtained from those in the case of spectrogram for the truncated Gaussian window $w_h(mT)$.

4) General Form of Variance for Small Signal Rate: Let us consider the value of $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ for an arbitrary quadratic TFD in the case of FM signals with a small signal rate within the kernel width $\phi^{(2)}(t) = a \to 0$. Then, $R_h(t)$, (22), and (23) takes the form

$$R_h(t) \cong B_h(0, 2) \to h^2 \cdot b_{0,2} \quad (48)$$

where $b_{0,2}$ is a constant depending on the kernel $\varphi(t, \tau)$. It is defined by (23). In this case, $C_f(0, 0; -\widetilde{\mathbf{\Psi}}_h^2)$ from (25) becomes

$$C_f(0, 0; -\widetilde{\mathbf{\Psi}}_h^2) \to |A(t)|^2 \cdot ThC_C \quad (49)$$

where C_C is the kernel $\varphi(t, \tau)$-dependent constant

$$C_C = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\widetilde{\varphi}(t_1, \tau)\widetilde{\varphi}^*(t_2, \tau)$$
$$\times(t_1 - \tau)(t_2 - \tau)dt_1 dt_2 d\tau. \quad (50)$$

Therefore, substituting (48) and (49) into (25), a general form [valid for any TFD and small $\phi^{(2)}(t) = a$] of $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ can be obtained as

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\} \cong \frac{\sigma_\varepsilon^2}{8\left|A(t)\right|^2}\cdot\frac{C_C}{\left|b_{0,2}\right|^2}\cdot\frac{T}{h^3}. \quad (51)$$

Note that (35) and (47) are just the special cases of (51).

Bias dependence on the kernel width h can easily be obtained from (30) as

$$E\{\triangle\widehat{\omega}_h(t)\} \cong C_B\phi^{(3)}(t)h^2 \tag{52}$$

where $b_{k,l}$ are defined by (23) and

$$C_B = \left(b_{2,2} + \frac{1}{3}b_{0,4}\right) / \left(2b_{0,2}\right). \tag{53}$$

This expression in not restricted to the small $\phi^{(2)}(t)$ since $\phi^{(2)}(t)$ is multiplied by higher order moments of the kernel, which are already small. The total IF estimation MSE as a function of h, for the case presented here, is

$$MSE = \mathrm{var}\{\triangle\widehat{\omega}_h(t)\} + (E\{\triangle\widehat{\omega}_h(t)\})^2$$

$$= \frac{\sigma_\varepsilon^2}{8\left|A(t)\right|^2}\cdot\frac{C_C}{\left|b_{0,2}\right|^2}\cdot\frac{T}{h^3} + (C_B\phi^{(3)}(t))^2\cdot h^4. \tag{54}$$

The MSE minimization, with respect to h, then reduces to the cases studied in [15], [16], and [27].

V. Multicomponent Signals

Consider a multicomponent signal

$$f(t) = \sum_{p=1}^{M} f_p(t) = \sum_{p=1}^{M} A_p e^{j\phi_p(t)} \tag{55}$$

and assume a definition of the IFs as the phase derivatives of the signal's components $\phi_p'(t) = d\phi_p(t)/dt$, $p = 1, 2, ..., M$. The IF estimation problem here consists of calculating the estimates of $\phi_p'(t)$, $p = 1, 2, ..., M$.

In this case, a quadratic TFD $C_f(t, \phi_i'(t); \widetilde{\mathbf{\Phi}}_h)$ from (24), at the frequency $\omega = \phi_i'(t)$, can be written as

$$C_f(t, \phi_i'(t); \widetilde{\mathbf{\Phi}}_h)$$

$$= \sum_{m_1=-\infty}^{\infty}\sum_{m_2=-\infty}^{\infty} \widetilde{\Phi}_h(m_1T, m_2T)$$

$$\times\left(\sum_{p=1}^{M}\sum_{q=1}^{M} f_p(t+m_1T)f_q^*(t+m_2T)\right)$$

$$\times e^{-j\phi_i'(t)(m_1-m_2)T}. \tag{56}$$

Based on the new kernel $\widetilde{\Phi}_h(m_1T, m_2T)$ definition, in the case of stationary, white, complex, Gaussian noise [see (76) from the Appendix], (56) can be written in the form

$$C_f(t, \phi_i'(t); \widetilde{\mathbf{\Phi}}_h)$$

$$= \sigma_\varepsilon^2 \sum_{n=-\infty}^{\infty}\sum_{p=1}^{M}\sum_{q=1}^{M}\Big[\sum_{m_1=-\infty}^{\infty} \widetilde{\varphi}_h(m_1T, nT)$$

$$\times(n-m_1)Tf_p(t+m_1T)e^{-j\phi_i'(t)m_1T}\Big]$$

$$\times\Big[\sum_{m_2=-\infty}^{\infty} \widetilde{\varphi}_h(m_2T, nT)$$

$$\times(n-m_2)Tf_q(t+m_2T)e^{-j\phi_i'(t)m_2T}\Big]^*. \tag{57}$$

The terms within summation could be considered to be the short-time Fourier transforms of the signal components $f_p(t)$ with the window function $w_\varphi(mT, nT) = \widetilde{\varphi}_h(mT, nT)(n-m)T$,

$$STFT_{f_p}(t, \phi_i'(t); nT)$$

$$= \sum_{m=-\infty}^{\infty} \widetilde{\varphi}_h(mT, nT)(n-m)T$$

$$\times f_p(t+mT)e^{-j\phi_i'(t)mT}. \tag{58}$$

Now, $C_f(t, \phi_i'(t); \widetilde{\mathbf{\Phi}}_h)$ can be written as

$$C_f(t, \phi_i'(t); \widetilde{\mathbf{\Phi}}_h)$$

$$= \sigma_\varepsilon^2 \sum_{n=-\infty}^{\infty}\left|\sum_{p=1}^{M} STFT_{f_p}(t, \phi_i'(t); nT)\right|^2. \tag{59}$$

Short-time Fourier transform $STFT_{f_p}(t, \omega; nT)$ is concentrated around the IF $\phi_p'(t)$. The concentration of $STFT_{f_p}(t, \omega; nT)$ is determined by the window $\widetilde{\varphi}_h(mT, nT)(n-m)T$ in (58). For example, in the pseudo WD case, the kernel function $\widetilde{\varphi}_h(mT, nT)(n-m)T$ is completely concentrated along $m = -n$ line, whereas in the cases of commonly used RID, the kernel functions $\widetilde{\varphi}_h(mT, nT)(n-m)T$ are spread around $m = -n$ line. In the spectrogram $\widetilde{\varphi}_h(mT, nT) = w_h(mT)w_h(nT)$, the window width is determined by $w_h(mT)$. This means that in the case of the spectrogram and the RID, when $STFT_{f_p}(t, \omega; nT)$

and $STFT_{f_q}(t,\omega;nT)$ do not overlap, for $p \neq q$, the following holds:

$$C_f(t,\phi_i'(t);\widetilde{\mathbf{\Phi}}_h) \cong \sigma_\varepsilon^2 \sum_{n=-\infty}^{\infty}\sum_{p=1}^{M}\left|STFT_{f_p}(t,\phi_i'(t);nT)\right|^2 \quad (60)$$

and

$$C_f(t,\phi_i'(t);\widetilde{\mathbf{\Phi}}_h) \cong \sigma_\varepsilon^2 \sum_{n=-\infty}^{\infty}\left|STFT_{f_i}(t,\phi_i'(t);nT)\right|^2 = C_{f_i}(t,\phi_i'(t);\widetilde{\mathbf{\Phi}}_h). \quad (61)$$

We can conclude that for the spectrogram and TFDs from the RID class, the IF estimation analysis derived for the case of monocomponent signal $f(t)$ remains valid for the multicomponent signals (55), for each component separately, when they are well separated in time-frequency plane.

In contrast to the RID, when the pseudo WD is used in the IF estimation, $STFT_{f_p}(t,\phi_i'(t);nT)$ has the form

$$STFT_{f_p}(t,\phi_i'(t);nT) = w_h^2(nT)(2nT)f_p(t-nT)e^{j\phi_i'(t)(nT)} \quad (62)$$

since $\widetilde{\varphi}_h(mT,nT) = w_h(mT)\delta(m+n)w_h(nT)$, and, consequently, in the case of nonoverlapping components $f_p(t)$, $p = 1,2,...,M$, (59) takes a simple form

$$C_f(t,\phi_i'(t);\widetilde{\mathbf{\Phi}}_h) = 4\sigma_\varepsilon^2\left(\sum_{p=1}^{M}A_p^2\right)\sum_{n=-\infty}^{\infty}w_h^4(nT)(nT)^2. \quad (63)$$

It can be concluded that variance (35) (derived for the case of monocomponent signals) for the estimation of $\phi_i'(t)$, by applying pseudo WD, increases with the number of components as

$$\mathrm{var}\{\triangle\widehat{\omega}_{h_i}(t)\} = \frac{\sigma_\varepsilon^2}{4A_i^4}(2\sum_{p=1}^{M}A_p^2+\sigma_\varepsilon^2)W_w\frac{T}{h^3}. \quad (64)$$

This is in accordance with the results from [9] for this special case.

The above conclusions for multicomponent signals will be illustrated and confirmed statistically in the next section.

VI. Numerical Implementation and Statistical Analysis

A. Monocomponent Signals

Obtained theoretical results for the variance of monocomponent signals (Proposition 1) are checked statistically and are presented in Fig. 1*a*)-*c*). The following quadratic TFDs are considered:

- pseudo WD with the Hanning window
$$\varphi(mT,nT) = w(nT)\delta(m)w(nT);$$
- spectrogram with the Hanning window
$$\varphi(mT,nT) = w(mT+nT)w(mT-nT);$$
- Born-Jordan distribution (BJD)
$$c(kT,2nT) = \frac{\sin((\pi/T)kTnT)}{(\pi/T)kTnT}$$
$$\varphi(mT,nT) = DFT_k\{c(kT,2nT)\};$$
- Choi-Williams distribution (CWD)
$$c(kT,2nT) = \exp(-((2\pi/T)kTnT/\sigma)^2),$$
$$\sigma = \sqrt{2}\pi$$
$$\varphi(mT,nT) = DFT_k\{c(kT,2nT)\}.$$

The linear FM signal $f(t) = \exp(j32\pi at^2)$ is assumed, within the time interval $t \in [-1,1]$ and with the sampling period $T = 1/128$. The IF of this signal varies from zero up to the maximal possible frequency (when $a = 1$), without aliasing, within the considered interval. In the case of spectrogram the signal $f(t)$ is sampled at the Nyquist rate $T = 1/64$. The signal $f(t)$ is corrupted by a stationary, Gaussian, white noise with variance $\sigma_\varepsilon = 0.25$. Analysis of the various σ_ε values (or SNR) influence on the IF estimation is performed in Section VI-D. Time-frequency plane consists of 256×256 points in all considered cases. The BJD and the CWD are calculated by using the kernel definition in (θ,τ) domain, and the signal's ambiguity function. Note that the results for the CWD highly depend on the parameter σ. Thus, any comparison is relative. Here, we have chosen the CWD parameter according to the results from [23]. The analysis of kernel width influence is done in Section VI-B. Note

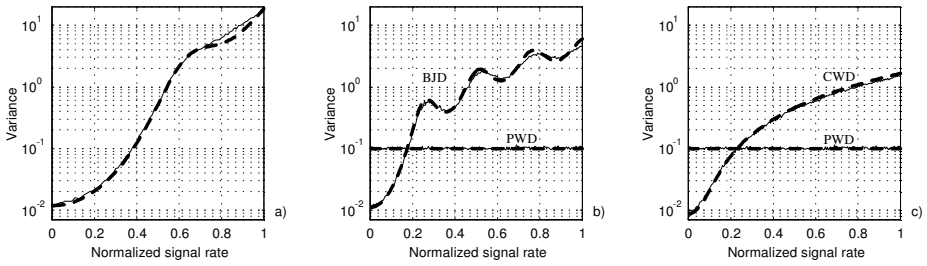

Fig. 1. IF variance obtained theoretically (dashed thick line), and statistically (thin solid line) as a function of the normalized values of signal rate $\phi^{(2)}(t) = a$; *a*) Spectrogram, *b*) Born-Jordan distribution, and pseudo Wigner distribution, *c*) Choi-Williams distribution and pseudo Wigner distribution. Note that $a = 0$ corresponds to the pure sinusoid, whereas the value of $a = 1$ corresponds to the diagonal of the considered time-frequency domain.

that for these parameters, the spectrogram behaves the best for small a ($a < 0.4$). For $a > 0.4$, the variance is smallest in the pseudo WD. Therefore, these two distributions could be reasonable choice for these values of a and monocomponent signals. However, for multicomponent signals, in general, the pseudo WD cannot be used, and we can see that for large a, the BJD and CWD outperform the spectrogram. Once more, we want to point out that any comparison is relative and would require the parameter optimization for each distribution and each signal before a general conclusion could be drawn.

A very high agreement of the theoretical results (dashed thick line) and the statistical data (thin solid line) can easily be observed in Fig. 1. Theoretical values are produced by applying the derived expressions (24) and (25), whereas the statistical data are obtained by running 128 simulations, i.e., in $128 \times 192 = 24576$ points. Typical error functions for one realization and three values of a are given in Fig. 2. Note that $\mathrm{var}\{\triangle\widehat{\omega}_h(t)\}$ in the BJD and the CWD cases increases (as in the case of spectrogram) as a increases. For small $a \to 0$, they have lower variance than the pseudo WD, whereas by increasing a, they perform worse than the pseudo WD. These conclusions are expected since the RID significantly reduce noise energy located far from the θ, τ axes. For the signals whose ambiguity function lies along the θ, τ axes (as in the case of linear FM signals with $a \to 0$), the RID do not degrade signal representation. On the other hand, for linear FM signals with larger values of a, the RID significantly degrade representation of the analyzed signal. Consequently, in this case, it may happen that the TFDs from RID class have worse performance than the pseudo WD. Oscillations in variance for the BJD case are due to its pseudo form used in numerical implementations. Namely, considering finite support of the BJD, significant kernel values can be truncated. Since they are oscillatory, they can cause variance oscillations as well. Note that the obtained variances for $a = 0$, in the case of spectrogram, with a Hanning window and the cases of considered RID (BJD and CWD) are almost equal. However, if a rectangular window were used in the spectrogram definition, its IF estimation variance would be 2.34 times smaller [according to (47) and the values of constant S_w given in the Table I].

B. Influence of the Kernel Width in Time-Lag Domain

The influence of kernel width is illustrated on the spectrogram, the CWD, and the pseudo WD. The IF variance decreases in the pseudo WD as h increases, as expected from (35) [see Fig. 3c)]. In addition, for the pseudo WD case, the IF variance does not depend on the signal rate a. For the IF estimation based on the spectrogram and the CWD, the variance is highly dependent on signal rate a. For small

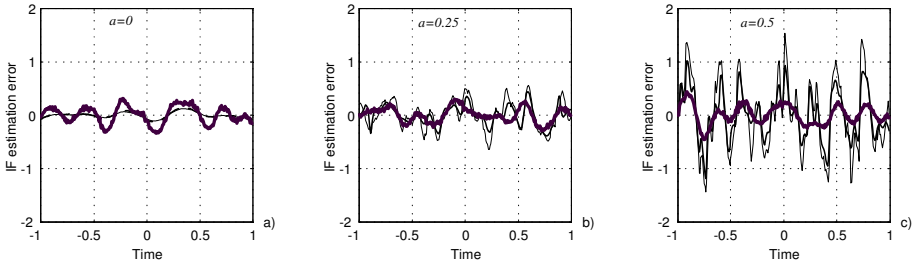

Fig. 2. Typical realization of the IF estimation error in the cases of the pseudo Wigner distribution (the thickest line), the Choi-Williams distribution (line), and the Born-Jordan distribution (the thinnest line), for various values of a: a) $a = 0$, b) $a = 0.25$, c) $a = 0.5$.

a, close to $a = 0$, an increase of the kernel width h significantly increases the signal energy, whereas the amount of noise (depending on the kernel energy) is only slightly increased. Thus, the improvement in the estimation (decrease of the IF estimation variance) is significant with larger h; see Fig. 3a) and b). Note that for small a, the spectrogram and the CWD produce better estimation results than the WD (calculated with a Hanning window). For very small values of a, the CWD (and other TFDs from the RID class) has the same behavior of the IF estimation variance as the pseudo WD [see Fig. 3a) and b) for $a = 0$ and Fig. 3c)], with different constant factors,

$$\mathrm{var}\{\triangle\widehat{\omega}_h(t)\} \sim \frac{1}{h^3}. \tag{65}$$

This relation is in accordance with (35), (47), and (51).

However, for high values of the signal rate a, the kernel width increase almost does not influence the signal energy, whereas it slightly increases the energy of noise. It results in the IF estimation variance increase [Fig. 3a) and b) for $a = 0.5$ and $a = 1$]. For these values of a, the spectrogram and the CWD produces worse results than the pseudo WD, as expected and shown in Section VI-A.

C. Multicomponent Signals

Theoretical results for the multicomponent signal (see Section V) are checked statistically on the spectrogram, BJD, CWD, and the pseudo WD. A sum of parallel linear FM noisy signals with the normalized amplitudes $A = 1$ is considered. In the IF estimation, the TFD maximum is detected, and then, a small region around the detected maximum is excluded. The IF of the next component is estimated as the position of the maximum in the remaining part of frequency axis.

For the spectrogram, we have obtained that the IF estimation variance depends neither on the number of component nor on the distance between components, as far as they do not overlap. Thus, all lines in Fig. 4a) overlap and coincide with the result for monocomponent signal from Fig. 1a). Similar situation takes place in the BJD and the CWD [see Fig. 4b) and c)]. When the pseudo WD is used in the IF estimation, then the variance in multicomponent signal is dependent both on the number and the total power of components. The results for monocomponent, two-component, and three-component signals are presented in Fig. 4d). The variance increases with the number of components, independently on the distance between components, exactly as expected from (64). In order to illustrate this theoretical result, the components are taken so that the cross-terms do not overlap with the auto-terms.

In Fig. 5, the IF estimation MSE ($MSE = \mathrm{var}\{\triangle\widehat{\omega}_h(t)\} + (E\{\triangle\widehat{\omega}_h(t)\})^2$) is presented. It can be seen that the MSE coincides with the IF estimation variance in all cases, except for the BJD and the CWD for close components and high signal rate a (close to the diagonal of the time-frequency plane $a = 1$); see the

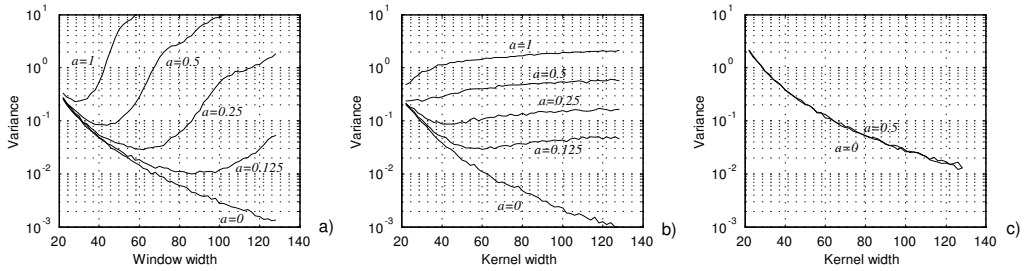

Fig. 3. IF variance as a function of the kernel width h for various values of the normalized signal rate a: a) Spectrogram, b) Choi-Williams distribution, c) pseudo Wigner distribution.

solid gray line in Fig. 5b) and c). Then, components exhibit a high mutual influence, and they start overlapping with non-negligible values, thus introducing the bias. The analytical treatment of this effect was very complex, and it is not given in the paper. For very small values of a, some increase in the MSE for two-component signal, with close components, can be spotted in the BJD and CWD as well. In this case, the variance is already very small, and the bias, although very small, is noticable due to the logarithmic scale. When the signal is monocomponent (thin solid line) or when the components in a multicomponent signal are well-separated (dashed line), the MSE coincides with the variance, meaning that the estimation can be treated as unbiased.

D. High Noise Effect in the IF Estimation

Sources of estimation error in the IF estimation based on the TFD maxima are

1) bias;

2) random deviation of the maxima within the auto-term caused by the small noise (This noise was the topic of this paper. It can cause that some of the auto-term points surpass the value of true maximum at the IF, producing the estimation error);

3) when the noise is very high, some TFD values that are outside the signal's auto-terms, can surpass all the auto-term values.

Then, a false maxima position is detected. It causes large random IF errors, uniformly distributed over the entire frequency range. When this kind of error starts to occur, it dominantly degrades the estimation performance. A detailed analysis of this kind of error is presented in [10]. Here, we will present the statistical data for the considered kernels, which are in full accordance with the analysis and results from [10]. The IF estimation variance in the BJD, the CWD, and the pseudo WD, as a function of the input SNR, is presented in Fig. 6. Bellow the SNR value, when the high errors appear, a significant increase in the IF estimation variance can be seen. For pseudo WD, this SNR is approximately equal to 4 dB, and unlike in the case of BJD and CWD, it is independent on the signal rate a.

A very simplified example of the high noise analysis in the pseudo WD will be given next. It is easy to conclude that the ratio of the pseudo WD maximal auto-term value (A_{WD}) and the standard deviation of the pseudo WD values (σ_{WD}) for a noisy linear FM signal is [2], [26], [28]

$$\frac{A_{WD}}{\sigma_{WD}} = \frac{A^2 \sum_{m=-N/2}^{N/2} w(m)w(-m)}{\sqrt{\sigma_\varepsilon^2(2A^2+\sigma_\varepsilon^2)\sum_{m=-N/2}^{N/2}\left(w(m)w(-m)\right)^2}} \tag{66}$$

where A is the amplitude of the input signal, and σ_ε^2 is the variance of input noise. Since the resulting noise in the pseudo WD may be considered as Gaussian [10], then a point outside the auto-term will have a value greater than $\kappa\sigma_{WD}$ with probability $P(\kappa)$. For example, for $\kappa = 3$, that probability is $P(3) =$

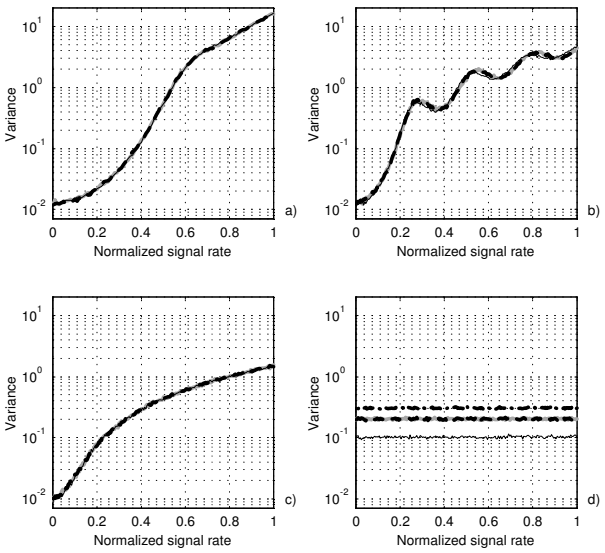

Fig. 4. IF variance as a function of the normalized signal rate a for monocomponent and multicomponent signals: a) Spectrogram, b) Born-Jordan distribution, c) Choi-Williams distribution, d) Pseudo Wigner distribution. Thin solid line is for monocomponent signals (as in Fig.1); dashed line is for a two-component signal with the distance between the components equal to one half of the whole frequency range; solid gray line is for a two-component signal with the distance between the components equal to one fourth of the whole frequency range; dot-dashed line in the pseudo Wigner distribution is for the three-component signal.

$(1 - 0.9973)/2 = 0.00135$, whereas for $\kappa = 6$, it is $P(6) = 10^{-6}$. Note that for each instant, there are slightly less than N points outside the auto-term.

For the example presented in Fig. 6, where the Hanning window $w(m)$ of $N = 64$ samples is used, with the normalized amplitude $A = 1$, we can conclude that the ratio $A_{WD}/\sigma_{WD} = 3$ (when $A_{WD} = \kappa\sigma_{WD}$) will cause one high error, very roughly speaking, in less than 12 instants [since $1/(12 \times 60) \sim P(3)$]. Since the value of this error can be very high, with this relatively high frequency of occurrence, it will be enough to degrade the IF estimation performance. For $A_{WD}/\sigma_{WD} = 6$, frequency of the high error occurrence is very low, even for a large number of points, since $P(6) = 10^{-6}$. Thus, the transition region (between no high errors and very frequent high errors) is approximately within $3 < A_{WD}/\sigma_{WD} < 6$. From (66), for the considered example, we get $3 < 5.78/\sqrt{\sigma_\varepsilon^2(2 + \sigma_\varepsilon^2)} < 6$, corresponding to

$$-0.7\,dB < SNR < 4.1\,dB$$

which is in agreement with Fig. 6c). Again, this is only a very rough analysis of the high noise influence. A rigorous analysis may be found in [10].

VII. Conclusion

In this paper, we have performed the IF estimation analysis based on the general quadratic shift-covariant class of TFDs. The exact bias and variance expressions are derived. It is shown that the IF estimation variance is closely related to the TFD of the non-noisy signal. The expressions, in the cases of most frequently used TFDs, are obtained as special cases of the general analysis. The results obtained for monocomponent signal remain valid for the multicomponent signals and TFDs from the RID class when the components are well separated. In the pseudo WD case, the estimation variance depends on the

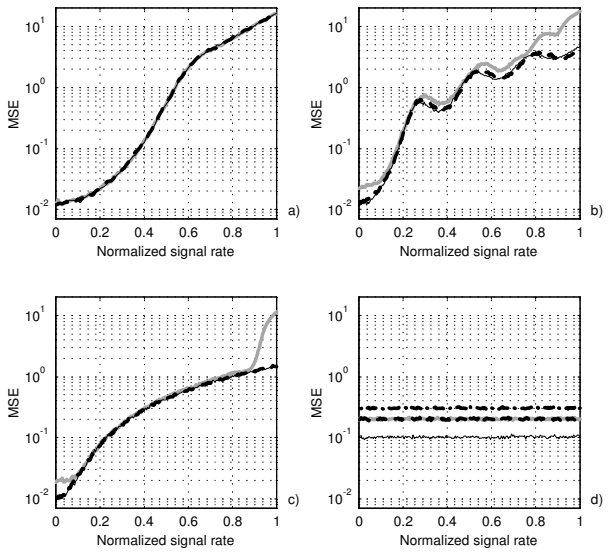

Fig. 5. MSE of the IF estimation (variance plus squared bias) as a function of the normalized signal rate a, for monocomponent and multicomponent signals: a) Spectrogram, b) Born-Jordan distribution, c) Choi-Williams distribution, d) Pseudo Wigner distribution. Thin solid line is for monocomponent signals (as in Fig.1); dashed line is for a two-component signal with the distance between the components equal to one half of the whole frequency range; solid gray line is for a two-component signal with the distance between the components equal to one fourth of the whole frequency range; dot-dashed line in the pseudo Wigner distribution is for the three-component signal.

total power of all signal components. The obtained results are checked numerically and statistically.

VIII. Appendix

Proof of Proposition 1: The general expression for the variance of the estimation error is given by (19), where the factor Q_h is defined by (14). According to the definition of Q_h, we can conclude that $E\{Q_h\} = 0$ since, in the case of white noise $\varepsilon(nT)$, $E\{\partial C_x(t,\omega;\varphi_h)/\partial\omega|_0\delta_\varepsilon\} = 0$, and $E\{\partial C_x(t,\omega;\varphi_h)/\partial\omega|_0\delta_{\varepsilon^2}\} = 0$. Therefore

$$\begin{aligned}\mathrm{var}\{Q_h\} &= E\{|Q_h|^2\}\\ &= E\Big\{\Big|\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon + \tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\Big|^2\Big\}.\end{aligned} \tag{67}$$

By applying the property that the odd-order moments of the zero mean $\varepsilon(nT)$ are equal to zero [20], it follows that

$$E\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon \times (\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2})^*\} = 0,$$

and

$$E\{(\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon)^* \times \tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\} = 0.$$

Thus, $\mathrm{var}\{Q_h\}$ may be written as

$$\begin{aligned}&\mathrm{var}\{Q_h\}\\ &= \mathrm{var}\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_\varepsilon\} + \mathrm{var}\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\}.\end{aligned} \tag{68}$$

The first term in (68) is highly signal and noise dependent. Second term is signal independent and time-frequency invariant for the case of stationary noise [14]-[16]. In the case of complex, Gaussian noise $\varepsilon(nT)$, [2], [11], [16], [26], [28], *the second term* from (68) can be written in the following form:

$$\begin{aligned}&\mathrm{var}\left\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\right\} = 4\\ &\times \sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}\sum_{m_1=-\infty}^{\infty}\sum_{m_2=-\infty}^{\infty}\end{aligned}$$

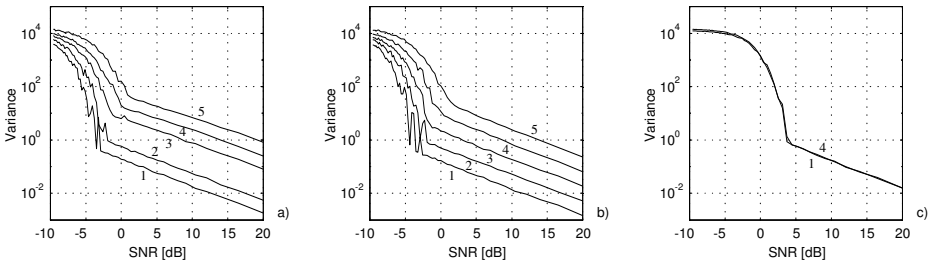

Fig. 6. IF estimation variance as a function of the input SNR, for various normalized signal rates: 1) $a = 0$, 2) $a = 0.125$, 3) $a = 0.25$, 4) $a = 0.5$, 5) $a = 1$. a) Born-Jordan distribution, b) Choi-Williams distribution, c) pseudo Wigner distribution.

$$\begin{aligned}
&\times\varphi_h(m_1T,n_1T)\varphi_h^*(m_2T,n_2T)\\
&\times[R_{\varepsilon\varepsilon}(t+m_1T+n_1T,t+m_1T-n_1T)\\
&\times R^*_{\varepsilon\varepsilon}(t+m_2T+n_2T,t+m_2T-n_2T)\\
&+R_{\varepsilon\varepsilon}(t+m_1T+n_1T,t+m_2T+n_2T)\\
&\times R^*_{\varepsilon\varepsilon}(t+m_1T-n_1T,t+m_2T-n_2T)]\\
&\times(n_1T)(n_2T)e^{-j2\phi'(t)(n_1-n_2)T} \qquad (69)
\end{aligned}$$

where $R_{\varepsilon\varepsilon}(t+mT,t+nT) = E\{\varepsilon(t+mT)\varepsilon^*(t+nT)\}$ is the noise $\varepsilon(nT)$ auto-correlation function.

Special Case: For *nonstationary, complex, white, Gaussian noise*, $R_{\varepsilon\varepsilon}(t+mT,t+nT) = I(t+mT)\delta(m-n)$, $I(t)\geq 0$, we get

$$\begin{aligned}
&\mathrm{var}\left\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\right\}\\
&=4\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}|\varphi_h(mT,nT)|^2(nT)^2\\
&\times I(t+mT+nT)I^*(t+mT-nT)\\
&=4C_I(t,0;|\varphi_{h_1}|^2) \qquad (70)
\end{aligned}$$

where $\varphi_{h_1}(mT,nT) = \varphi_h(mT,nT)(nT)$. Thus, in this case, the noise-only dependent part of variance may be represented as a quadratic TFD of $I(t)$, with the new kernel $|\varphi_{h_1}(mT,nT)|^2$, [24].

Special Case: For *stationary, complex, white, Gaussian noise*, $I(t) = \sigma_\varepsilon^2$, we have

$$\begin{aligned}
&\mathrm{var}\left\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon^2}\right\}\\
&=4\sigma_\varepsilon^4\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}|\varphi_h(mT,nT)|^2(nT)^2\\
&=4\sigma_\varepsilon^4 W_h. \qquad (71)
\end{aligned}$$

Note that, as $T\to 0$, W_h is reduced to

$$W_h\to T^2W = T^2\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}|\varphi(t,\tau)|^2\tau^2dtd\tau \qquad (72)$$

where W depends on the kernel $\varphi(t,\tau)$ type only.

The first term from (68), for real and symmetric kernel $\varphi_h(mT,nT)$, and in the case of complex, Gaussian noise $\varepsilon(nT)$ may be represented as

$$\begin{aligned}
&\mathrm{var}\left\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon}\right\}\\
&=\sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}\sum_{m_1=-\infty}^{\infty}\sum_{m_2=-\infty}^{\infty}\\
&\times\widetilde{\varphi}_h(m_1T,n_1T)\widetilde{\varphi}_h^*(m_2T,n_2T)\\
&[f(t+m_1T)f^*(t+m_2T)R^*_{\varepsilon\varepsilon}(t+n_1T,t+n_2T)\\
&+f^*(t+n_1T)f(t+n_2T)R_{\varepsilon\varepsilon}(t+m_1T,t+m_2T)]\\
&\times(n_1-m_1)(n_2-m_2)T^2\\
&\times e^{-j\omega(m_1-m_2)T}e^{-j\omega(n_2-n_1)T}|_0. \qquad (73)
\end{aligned}$$

Applying $\widetilde{\varphi}_h(m_1T,nT) = \widetilde{\varphi}_h(nT,m_1T)$, and $R_{\varepsilon\varepsilon}(t+mT,t+nT) = I(t+mT)R_{\varepsilon\varepsilon}(m-n)$, $I(t)\geq 0$, we get

$$\mathrm{var}\left\{\tfrac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0\delta_{\varepsilon}\right\}$$

$$= 2 \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \widetilde{\Phi}_h(m_1T, m_2T) \times[f(t+m_1T)e^{-j\phi'(t)m_1T}] \times[f(t+m_2T)e^{-j\phi'(t)m_2T}]^* = 2C_f(t, \phi'(t), \widetilde{\Phi}_h) \quad (74)$$

where $C_f(t, \phi'(t), \widetilde{\Phi}_h)$ is a quadratic TFD [with the new kernel $\widetilde{\Phi}_h(m_1T, m_2T) = \Phi_h((m_1+m_2)T/2, (m_1-m_2)T/2)$] of the analyzed signal $f(t)$ at the signal's IF $\phi'(t)$. The general form of the new kernel $\widetilde{\Phi}_h(m_1T, m_2T)$ is [24]

$$\widetilde{\Phi}_h(m_1T, m_2T) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \widetilde{\varphi}_h(m_1T, n_1T) \times\widetilde{\varphi}_h^*(m_2T, n_2T)(n_1-m_1)T(n_2-m_2)T \times I(t+n_2T)R_{\varepsilon\varepsilon}(n_2-n_1)e^{-j\phi'(t)(n_2-n_1)T}. \quad (75)$$

Special Case: For *stationary, complex, white, Gaussian noise*, we get

$$\widetilde{\Phi}_h(m_1T, m_2T) = \sigma_\varepsilon^2 \sum_{n=-\infty}^{\infty} \widetilde{\varphi}_h(m_1T, nT)\widetilde{\varphi}_h^*(m_2T, nT) \times(n-m_1)(n-m_2)T^2. \quad (76)$$

For finite limits, this is a matrix multiplication form

$$\widetilde{\mathbf{\Phi}}_h = \sigma_\varepsilon^2[\mathbf{A}_{n-m}. * \widetilde{\boldsymbol{\varphi}}_h] \times [\mathbf{A}_{m-n}. * \widetilde{\boldsymbol{\varphi}}_h] \quad (77)$$

where $\mathbf{A}_{n-m}$ is a matrix with elements $A(m,n) = n-m$, for $m, n = 1, 2, ..., N$. Elements of matrix $\widetilde{\boldsymbol{\varphi}}_h$ are $\widetilde{\varphi}_h(mT, nT)$. Let us now introduce $\widetilde{\mathbf{\Psi}}_h = \mathbf{A}_{n-m}. * \widetilde{\boldsymbol{\varphi}}_h$. Then, because of symmetry and realness of matrix $\widetilde{\boldsymbol{\varphi}}_h$, $\widetilde{\varphi}_h^*(m_2T, nT) = \widetilde{\varphi}_h(nT, m_2T)$, and the antisymmetry of matrix $\mathbf{A}_{n-m}$, $\mathbf{A}_{n-m} = -\mathbf{A}_{m-n}$, we have

$$\widetilde{\mathbf{\Phi}}_h = -\sigma_\varepsilon^2 \widetilde{\mathbf{\Psi}}_h^2. \quad (78)$$

Thus

$$\mathrm{var}\left\{ \frac{\partial C_x(t,\omega;\varphi_h)}{\partial\omega}|_0 \delta_\varepsilon \right\} = 2\sigma_\varepsilon^2 C_f(t, \phi'(t); -\widetilde{\mathbf{\Psi}}_h^2). \quad (79)$$

Special Case: For *nonstationary, complex, white, Gaussian noise*, we have

$$\widetilde{\Phi}_h(m_1T, m_2T) = \sum_{n=-\infty}^{\infty} \widetilde{\varphi}_h(m_1T, nT)\widetilde{\varphi}_h^*(m_2T, nT) \times(n-m_1)(n-m_2)T^2 I(t+nT) = -\widetilde{\mathbf{\Psi}}_h \mathbf{I}_t \widetilde{\mathbf{\Psi}}_h^* \quad (80)$$

where matrix $\mathbf{I}_t$ is defined in Proposition 1. Substitution of (70) and (80) into (19) gives (24) of Proposition 1. In addition, by substituting (71) and (79) into (19), we get (25) of the special case of Proposition 1. ■

IX. Acknowledgment

The authors are thankful to the reviewers whose comments have helped to significantly improve the content and widen the scope of the results. This work is supported by the Volkswagen Stiftung, Federal Republic of Germany.

References

[1] M.G. Amin: "Spectral decomposition of time-frequency distribution kernels", *IEEE Trans. on Signal Processing,* vol.42, no.5, May 1994, pp.1156-1165.

[2] M.G. Amin: "Minimum-variance time-frequency distribution kernels for signals in additive noise", *IEEE Trans. on Signal Processing*, vol.44, no.9, Sept.1996, pp.2352-2356.

[3] B. Boashash: "Estimating and interpreting the instantaneous frequency of a signal - Part 1: Fundamentals", *Proc. IEEE*, vol.80, Apr.1992, pp.519-538.

[4] L. Cohen: "Time-frequency distributions - a review", *Proc. IEEE*, vol.77, July 1989, pp.941-981.

[5] L. Cohen: "Distributions concentrated along the instantaneous frequency", *SPIE, Advanced Signal Processing Alg., Architect., and Implementations*, vol.1348, 1990, pp.149-157.

[6] L. Cohen, *Time-frequency analysis*, Prentice Hall, Englewood Cliffs, N.J., 1995.

[7] L. Cohen, C. Lee: "Instantaneous bandwidth", in *Time-frequency signal analysis*, B. Boashash ed., Longman Cheshire, 1992.

[8] G.S. Cunningham, W.J. Williams: "Kernel decomposition of time-frequency distributions", *IEEE Trans. on Signal Processing*, vol.42, June 1994, pp.1425-1441.

[9] I. Djurović, LJ. Stanković: "A virtual instrument for time-frequency analysis", *IEEE Trans. on Instrumentation an Measurements*, vol.48, no.6, Dec.1999, pp.1086-1092.

[10] I. Djurović, LJ. Stanković: "Influence of high noise on the instantaneous frequency estimation

using time-frequency distributions", *IEEE Signal Processing Letters*, vol.7, no.11, Nov.2000, pp.317-319.

[11] S.B. Hearon, M.G. Amin: "Minimum-variance time-frequency distribution kernels", *IEEE Trans. on Signal Processing*, vol.43, no.5, May 1995, pp.1258-1262.

[12] F. Hlawatsch, G.F. Boudreaux-Bartels: "Linear and quadratic time-frequency signal representation", *IEEE Signal Processing Magazine*, Apr.1992, pp.21-67.

[13] J. Jeong, W.J. Williams: "Kernel design for reduced interference distributions", *IEEE Trans. on Signal Processing*, vol.40, no.2, Feb.1992, pp.402-412.

[14] V. Katkovnik: "Nonparametric estimation of instantaneous frequency", *IEEE Trans. on Information Theory*, vol.43, no.1, Jan.1997, pp.183-189.

[15] V. Katkovnik, LJ. Stanković: "Instantaneous frequency estimation using the periodogram with time-varying window length", *Signal Processing*, vol.67, no.3, June 1998, pp.345-358.

[16] V. Katkovnik, LJ. Stanković: "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length", *IEEE Trans. on Signal Processing*, vol.46, no.9, Sept.1998, pp.2315-2326.

[17] S.M. Kay: "Statistically/computation-ally efficient frequency estimation", *Proc. IEEE ICASSP*, New York, 1988, pp.2292-2295.

[18] B.C. Lovell, R.C. Williamson, B. Boashash: "The relationship between instantaneous frequency and time-frequency representations", *IEEE Trans. on Signal Processing*, vol.41, no.3, Mar.1993, pp.1458-1461.

[19] B.C. Lovell, R.C. Williamson: "The statistical performance of some instantaneous frequency estimators", *IEEE Trans. on Signal Processing*, vol.40, no.7, July 1992, pp.1707-1723.

[20] A. Papoulis, *Signal analysis*, McGraw-Hill Company, 1977.

[21] B.G. Quin: "Estimating frequency by interpolation using Fourier coefficients", *IEEE Trans. on Signal Processing*, vol.42, no.5, May 1994, pp.1264-1268.

[22] P. Rao, F.J. Taylor: "Estimation of the instantaneous frequency using the discrete Wigner distribution", *Electronics Letters*, vol.26, 1990, pp.246-248.

[23] LJ. Stanković: "Auto-term representation by the reduced interference distributions; A procedure for kernel design", *IEEE Trans. on Signal Processing*, vol.44, no.6, June 1996, pp.1557-1564.

[24] LJ. Stanković: "An analysis of noise in time-frequency distributions", *IEEE Signal Processong Letters*, vol.9, Sept.2002, pp.286-289.

[25] LJ. Stanković, M. Daković, V. Ivanović: "Performance of spectrogram as IF estimator", *Electronics Letters*, vol.37, no.12, June 2001, pp.797-799.

[26] LJ. Stanković, V. Ivanović: "Further results on the minimum variance time-frequency distribution kernels", *IEEE Trans. on Signal Processing*, vol.45, no.6, June 1997, pp.1650-1655.

[27] LJ. Stanković, V. Katkovnik: "Algorithm for the instantaneous frequency estimation using time-frequency distributions with adaptive window length", *IEEE Signal Processing Letters*, vol.5, no.9, Sept.1998, pp.224-227.

[28] LJ. Stanković, S. Stanković: "On the Wigner distribution of discrete-time noisy signals with application to the study of quantization effects", *IEEE Trans. on Signal Processing*, vol.42, no.7, July 1994, pp.1863-1867.

[29] K.M. Wong: "Estimation of the time-varying frequency of a signal: The Cramer-Rao bound and the application of Wigner distributions", *IEEE Trans. on Acoust., Speech, Signal Processing*, vol.38, 3, 1990, pp.519-535.

[30] D. Wu, J.M. Morris: "Discrete Cohen's class of distributions", in *Proc. of the IEEE Symposium of TFTSA*, Philadelphia, PA, Oct.1994, pp.532-535

[31] Y. Zhao, L.E. Atlas, R.J. Marks, II: "The use of cone-shaped kernels for generalized time-frequency representations on non-stationary signals", *IEEE Trans. on Acoust, Speech, Signal Processing*, vol.38, no.7, July 1990, pp.1084-1091.